J Financ Serv Res (2017) 52:277–305
DOI 10.1007/s10693-016-0258-x

Do Stars Shine? Comparing the Performance Persistence of Star Sell-Side Analysts Listed by Institutional Investor, the Wall Street Journal, and StarMine

Yury O. Kucheev[1,2,3] · **Felipe Ruiz**[1] · **Tomas Sorensson**[2,3]

Received: 1 June 2015 / Revised: 14 July 2016 / Accepted: 27 July 2016 /
Published online: 3 September 2016

Abstract We investigate the profitability persistence of the investment recommendations from analysts listed in four different star rankings, *Institutional Investor* magazine, *StarMine's* "Top Earnings Estimators" and "Top Stock Pickers", and *The Wall Street Journal*, and show the predictive power of each evaluation methodology. We found that only Buy and Strong Buy recommendations from the entire group of Star analysts outperform those of the Non-Stars in the year after election, while Sell and Strong Sell recommendations performed as those of the Non-Stars. We document that the highest average monthly abnormal return of holding a long-short portfolio, 0.97 %, is obtained by following the recommendations of the group of star sell-side analysts rated by *StarMine's* "Top Earnings Estimators" during the period from 2003 to 2014. Since earnings are one of the main drivers of stock prices, the results obtained are in line with the notion that focusing on superior earnings forecasts is one of the top requirements for successful stock picks.

Keywords Star analysts · Analyst recommendations · StarMine · Institutional Investor · The Wall Street Journal

JEL Classification G23 · G24

✉ Tomas Sorensson
tomas.sorensson@indek.kth.se

[1] Department of Industrial Management, Business Administration and Statistics, School of Industrial Engineering, Technical University of Madrid (UPM)/Universidad Politécnica de Madrid, c/ José Gutiérrez Abascal, 2, 28006 Madrid, Spain

[2] Department of Industrial Economics and Management, School of Industrial Engineering and Management, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden

[3] Swedish House of Finance, Stockholm School of Economics, Stockholm, Sweden

1 Introduction

This study analyzes whether investors can profit from the recommendations of ranked security analysts. We examine whether an investor's choice of a rating "agency" matters and how the methodologies used by different star rankings are able to predict the investment value of the recommendations. By investigating the precision of signals that the various methodologies use in determining who the stars are, we distinguish between those star-selection methodologies that capture a short-term stock-picking profitability and those methodologies that emphasize more persistent skills of the analysts. As a result, this study documents that there are star-selection methods that select analysts on the basis of more enduring analyst skills, and, thus, their stars' performance persists even after ranking announcement.

We compare the performance of the rankings by *The Wall Street Journal* and *StarMine* that are explicitly based on the analysts' past performance, which is *objectively* measured, with those of the *Institutional Investor* rankings that are based on *subjective* survey assessments by the analysts' buy-side clients. The expected differences between objective and subjective rankings in predicting stock recommendation performance depend on the persistence of the outperformance of a small group of objectively ranked analysts. Such outperformance is composed of both the stock-picking skill and luck. If outperformance is persistent, objective methods will have an edge over subjective methods. However, if outperformance is primarily due to luck and it is not persistent, then subjective rankings might work better if the buy-side has insight on which analysts have better stock-picking skills amidst the noise. We have to acknowledge that the different rankings could be directed to different clienteles rather than focusing on stock-picking skills. In contrast, our study focuses on the persistence in economic performance by the analysts, measured by portfolio returns.

Academic theory and banks do not reach the same conclusions about the value of security analysts. The semi-strong form of market efficiency states that investors should not be able to earn excess returns from trading on publicly available information, such as analysts' recommendations. However, banks and other firms spend large amounts of money on research departments and security analysts, presumably because they and their clients believe that security analysis can generate large abnormal returns. The importance of security analysis and analysts is also manifest in the establishment, in 1998, and growth of *StarMine*, a competitor to *The Wall Street Journal* and *Institutional Investor*'s rankings of analysts.[1] In the following, we analyze the economic value of security analysis – an activity performed by thousands of professionals in the finance industry with the goal of improving their clients' return performance.

The possibility that there could be profitable investment strategies based on the published recommendations of security analysts is supported by multiple studies (Stickel 1995; Womack 1996; Barber et al. 2001; Boni and Womack 2006; Barber et al. 2010; Loh 2010) that show that favorable (unfavorable) changes in individual analysts' recommendations are accompanied by positive (negative) returns at the time of their announcements. Hence, early work by Womack (1996) documents a post-recommendation stock price drift for upgrades that lasts up to one month and for downgrades that lasts up to six months.

Our perspective, however, differs from that of the above-mentioned studies. While the studies cited focus on measuring the average price reaction to changes in individual analysts' recommendations, we compare the profitability of recommendations issued by different groups

[1] *StarMine* states in a sales message on their homepage: "*StarMine is the world's largest and most trusted source of objective equity research performance ratings*" (StarMine 2015a).

of analysts. However, we pursue a common goal of providing evidence as to whether, assuming no transaction costs, profitable investment strategies could potentially be based on the use of analysts' recommendations. Specifically, we focus on differences between the rankings of security analysts by *Institutional Investor, The Wall Street Journal* and *StarMine* and on the profitability of their recommendations. Using this approach, we can determine whether investors can earn positive abnormal returns on the investigated portfolio strategies and whether differences in profitability are associated with the use of different star rankings. We also compare star analysts' recommendations with those of non-star analysts.

We use data from the Thomson Financials Institutional Brokers' Estimate System (I/B/E/S) *Detail Recommendations File* for the period from 2002 to 2013. We manually collected lists of star analysts from *Institutional Investor* magazine (October 2003 – October 2013*), The Wall Street Journal* (May 2003 – April 2013), and *StarMine* (October 2003 – August 2013). The lists of stars are matched with I/B/E/S by analysts' names and broker affiliations. Our final database contains 172,525 recommendations for 6443 companies listed on the NYSE, AMEX and NASDAQ markets that were announced between January 2002 and October 2014. The hand-collected database enables us to conduct original research by comparing the profitability of *Institutional Investor*'s rankings of analysts with the rankings of *StarMine* and *The Wall Street Journal*.

Using this database, we measure and compare the investment values of portfolios formed by recommendations of an entire group of star analysts (referred to as *Stars*[2]), a group of non-star analysts (*Non-Stars*), and groups of stars as indicated by the different rankings (*Institutional Investor* magazine, *StarMine's* "Top Earnings Estimators" and "Top Stock Pickers", and *The Wall Street Journal*). We divide our sample into two time frames, *Year Before*[3] and *Year After*, which correspond to the evaluation year and the one-year period after a particular star ranking is announced until the next announcement date (and twelve-month period for the year 2013 which was the last year for rankings lists in our dataset), respectively.

In line with Emery and Li (2009) and Fang and Yasuda (2014), we sort analysts according to their star/non-star status and use a well-established methodology of constructing dynamic buy-and-hold portfolios with a holding period of one year to form a "Long", "Hold", and "Short" portfolios for each group of analysts. The portfolio composition is formed according to the recommendations issued by a particular group of analysts. A Long portfolio includes all Buy and Strong Buy recommendations. Hold portfolio includes all Hold recommendations, while a Short portfolio contains Sell and Strong Sell recommendations. Each time an analyst reports that he or she has started covering a firm or changes his or her recommendation for a firm, the firm is included or excluded from the portfolio at the close of the recommendation announcement day (or at the close of the next trading day if the recommendation is issued after the closing of trading or on a non-trading day). Any returns that investors might have earned from prior knowledge of recommendations or from trading the recommended stocks during the recommendation day are not included in the return calculations. Time series of daily returns were used to estimate average risk-adjusted daily alphas for each portfolio. We then present the results as monthly abnormal returns in percent by multiplying daily values by 21 trading days.

[2] Throughout the paper, we capitalize and italicize the name of the group of *Stars* that consists of non-repeating names of star analysts from all four different rankings. All other stars have lower-case spelling. Particularly, lower-case spelling is used for either one particular group of star analysts or for the number-one ranked stars.

[3] By using *Year Before* comparison, we compare the performance of analysts during the evaluation year in a uniform way. We expect our result for the rankings to diverge due to the differences in election methods which are used by different rankings.

For our sample period, we find that only Strong Buy and Buy recommendations of star analysts generated higher monthly average excess returns (alphas) (0.33 %) than those recommendations by *Non-Stars* (0.18 %), while Hold, Sell and Strong Sell recommendations performed insignificantly different from those of the *Non-Stars*.

Among the entire groups of stars, the best performance was observed for *StarMine*'s "Top Earnings Estimators", with a monthly excess return of 0.97 % followed by *The Wall Street Journal* with 0.63 %, and *StarMine*'s "Top Stock Pickers" stars with 0.54 %. The worst performance was observed for *Institutional Investor*, with an excess return of 0.14 %. However, on a detailed level, Long portfolio of *StarMine*'s "Top Stock Pickers" is the number one portfolio, but its Short portfolio is the number three portfolio, which we interpret as suggesting that *StarMine*'s "Top Stock Pickers" might focus more on buy recommendations. Comparing the Long portfolios of the top-ranked analysts, we find that the analyst ranked number one by *StarMine*'s "Top Earnings Estimators" had higher returns than the group of *Non-Star* analysts and that the difference in returns was statistically significant.

Our results show that star analysts who are ranked in terms of the accuracy and timing of their earnings forecasts, as in the methodology of *StarMine*'s "Top Earnings Estimators", show performance that is more persistent from the year of evaluation to the year after than the star analysts who are ranked exclusively based on the previous performance of their recommendations (stars listed by *StarMine*'s "Top Stock Picker" and *The Wall Street Journal*). This result reveals that focusing on earnings forecasts when ranking analysts provides higher predictive power in selecting skilled analysts, while considering only the profitability of the previous year's recommendations leads to a large influence of luck. Previously, this result has also been documented by Loh and Mian (2006). We interpret our results as indirect empirical evidence in support for valuation models in the accounting and finance literature, underlining the role of future earnings in forecasting future stock price movements as in Ohlson (1995).

Our contribution is the comparison of four different star rankings with a focus on the profitability of investment recommendations using a recent dataset with a unique (hand-collected) list of star analysts. Emery and Li (2009) use the information ratio, which is the *t*-statistic of the intercept of the regression estimation, rather than a direct performance measure of the profitability of recommendations, as is used here. Fang and Yasuda (2014) use a 1994–2009 sample period and a 30-day holding period and find that recommendations issued by *Institutional Investor* stars have significantly higher returns than those of all other analysts (*Non-Stars*). In our study, we use a 2003–2014 sample period and a one-year time horizon for recommendations issued by analysts and find that *Institutional Investor* has no statistically significant predictive power. We document that other rankings using other measures in the evaluation of analysts give better predictive power.

1.1 Ranking evaluation approaches

Analysts are rated as a "star" based on their quality of previous reports, accuracy of forecasts and returns generated for clients (Loh and Mian 2006). The four different rankings of sell-side analysts in our study (see Table 1) can be divided into two main groups according to the evaluation approach used: objective (*StarMine* and *The Wall Street Journal*) or subjective (*Institutional Investor*). Two objective rankings are based exclusively on the investment value of recommendations: "Best on the Street," issued by *The Wall Street Journal*, and "Top Stock Pickers," issued by Thomson Reuters' *StarMine*. A third objective ranking is "Top Earnings Estimators," also issued by *StarMine*. It measures the accuracy and timing of each analyst's

Table 1 Description of rankings

Ranking name	Rating agency	Abbreviation used in this paper	Type of ranking	Measure	Measurement	Number of analysts per industry	Publication Month
"All-America Research Team"	Institutional Investor	I/I	Subjective / Qualitative	12 Criteria (most important: industry knowledge and integrity; least important: stock picking, and accuracy of EPS)	Survey	3 + Runners-up	October
"Top Earnings Estimators"	StarMine	STM-TEE	Objective / Quantitative	Accuracy and timing of earnings estimations	Calculation - EPS	3	October[a]
"Top Stock Pickers"	StarMine	STM-TSP	Objective / Quantitative	Excess returns on individual portfolios	Calculation - Recommendations	3	October[a]
"Best on the Street"	The Wall Street Journal	WSJ	Objective / Quantitative	Total score for stock returns	Calculation - Recommendations	5 in 2003–2011, 3 in 2012, 2013	May

[a] except of those lists of STM-TSP and STM-TEE that were announced in December 2009, May 2012, and August 2013

earnings forecast. A subjective ranking that uses mixed evaluation methods is the survey-based "All-America Research Team" issued by *Institutional Investor* magazine.

To select the members of the "All-America Research Team" ranking, *Institutional Investor* (I/I) magazine sends a questionnaire to buy-side investment managers asking them to evaluate various attributes of sell-side analysts. I/I magazine ranks three analysts in each industry and also provides names of so-called "runners-up" who are promising and could possibly be chosen in subsequent years. This list of stars is published in October and is usually supplemented by the 12 attributes in the survey that investors view as the most important to possess. Attributes such as industry knowledge and integrity are listed among the most important, while stock selection and earnings estimates are among the lowest-ranked attributes. Thus, the I/I ranking is not primarily focused on stock-picking ability but rather covers a wide range of attributes that are perceived to relate directly or indirectly to the ability of an analyst to make profitable recommendations. These attributes could be of high value for some clientele, though these qualitative attributes are not possible to measure by portfolio returns.

Previous research shows mixed results regarding the profitability of recommendations issued by I/I stars. Measuring the investment value of recommendations during the period from 1994 to 2009, Fang and Yasuda (2014) show that I/I stars outperformed the group of *Non-Stars*, finding Carhart 4-factor monthly alphas of 1.25 % for Long portfolios and −0.83 % monthly alphas for Short portfolios of I/I stars compared with 1.09 % and −0.71 % for Long and Short portfolios for *Non-Stars*, respectively. Using historical data from 1993 to 2005, Emery and Li (2009) investigate I/I and WSJ rankings. The authors identify the determinants of star status and compare the two rankings on the basis of earnings per share (EPS) accuracy and the industry-adjusted performance of investment recommendations in the year before and one year after analysts become stars. Emery and Li (2009) find, for the period from 1993 to 2005, that after becoming stars, star analysts' forecast accuracy of EPS does not differ from that of their non-star peers; the recommendations of I/I stars are not statistically better than those of *Non-Stars*, while the recommendations of WSJ stars are significantly worse. They conclude that both rankings are largely "popularity contests" and do not provide any significant investment value. In contrast, Leone and Wu (2007) investigate the investment value of I/I stars' recommendations issued from 1991 to 2000 and find that star analysts persistently issued profitable recommendations and that this outperformance was not due to luck but to a superior ability to pick stocks.

Since 1993, *The Wall Street Journal* (WSJ) has published a list of "Best on the Street" analysts (before 2000, this ranking was named "All-Star Analysts"), with five analysts ranked in each industry during 2003–2011 and three analysts per industry in the years of 2012 and 2013. This ranking is based on the score that an analyst obtained during the previous year, calculated as the sum of one-day returns of recommendations (if an investor would invest one day before a recommendation is announced and realize the return by the end of the recommendation day) (Emery and Li 2009). Such an evaluation methodology focuses on short-term price forecasts and favors analysts who issue recommendations on days when a price changes the most. At the same time, it penalizes analysts who issue their recommendations before or after such days of sharp price changes. Additionally, to benefit fully from such recommendations, investors should be able to receive a recommendation one day before it is announced, which could be the case for a limited number of investors with privileged access to analysts' recommendations. According to Yaros and Imielinski (2013), WSJ's evaluation method is not able to avoid analysts who announce their recommendations on the same day but after a significant price change has already occurred. All of these considerations may generate significant randomness in the election of analysts into the WSJ star ranking. As mentioned earlier, Emery and Li (2009) find that, after becoming stars, WSJ star

analysts issue recommendations that underperform the group of *Non-Stars*. They interpret this result as an effect of regression to the mean, as the short-term recommendation performance includes a substantial random component.

Thomson Reuters' *StarMine* "Top Stock Pickers" (STM-TSP) and "Top Earnings Estimator" (STM-TEE), which both include three analysts per industry, have been issued annually since 1998. They are both issued around October each year (except of those lists that were announced in December 2009, May 2012, and August 2013). The STM-TSP ranking is based on the excess returns of a non-leveraged portfolio built on all of the recommendations of each analyst. The returns of each analyst are calculated using the long and short buy-and-hold portfolio method relative to the market capitalization-weighted portfolio of all of the stocks in a given industry. The portfolio is rebalanced each month and whenever the analyst changes rating, adds coverage or drops coverage. The STM-TEE ranking measures the accuracy of each analyst's earnings forecasts and it is a measure of relative accuracy, since the analysts are compared against their peers. The measure accounts for several factors: the analyst's forecast error, the variance of the analysts' errors, the analyst's error compared to other analysts, the timing of the estimates, and the absolute value of the actual earnings of the firm. The measure is computed daily and aggregated to provide scores on individual stocks, industries and the analyst overall (StarMine 2015b). Up to 2012, STM-TEE's evaluation was based on earnings forecasts from the previous calendar-year. However, from 2012, STM-TEE uses earnings from the immediate year before announcement of the rankings lists. To summarize, the STM-TEE ranking differs from the STM-TSP and WSJ rankings since it does not consider the investment value of analysts' recommendations, thus STM-TEE does not measure the abnormal returns on portfolios.

Although *StarMine's* rankings appeared much later, they play an essential role in sell-side research by providing an "…influential and an important reference in the industry" (Kim and Zapatero 2011). According to Beyer and Guttman (2011) and Ertimur et al. (2011), many Wall Street firms use *StarMine* rankings when defining payments to their analysts. Recent work by Kerl and Ohlert (2015) investigates the accuracy of earnings per share forecasts and target prices of *StarMine* analysts compared with their non-star peers one year after the analysts became stars. They find that analysts possess a persistent ability to issue accurate earnings forecasts, and after becoming stars, they continue to issue more accurate earnings forecasts than non-star analysts. Regarding the accuracy of target prices (TP), the authors cannot find any difference between the two groups of analysts. The insignificant difference in TP forecasts could be due to the research methodology: star analysts with "Stock Picking Awards" and "Earnings Estimate Awards" are grouped together to compare their accuracy with that of *Non-Stars* without splitting the sample of *StarMine*'s stars into Top Stock Pickers and Top Earnings Estimators. However, according to the *StarMine* methodology for determining the "Stock Picking Awards", analysts are not evaluated on the basis of accuracy of EPS. Thus, it is possible that, even in the year before they receive an award, this mixed group of stars does not outperform non-star peers in terms of the accuracy of their forecasts. Furthermore, Kerl and Ohlert (2015) focus solely on the accuracy of EPS and TP and the factors that influence such accuracy and do not compare the performance of recommendations issued by star analysts with that of *Non-Stars*.

2 Data and descriptive statistics

We use four data sources. The Thomson Financials Institutional Brokers' Estimate System (I/B/E/S) *Detail Recommendations File* provides standardized stock recommendations for all

of the various brokers' scales by mapping all of the recommendations on a final scale from 1 to 5, where 1 corresponds to "Strong Buy", 2 to "Buy", 3 to "Hold", 4 to "Sell" and 5 to "Strong Sell". The *Daily Stock File* from the Center for Research in Security Prices (CRSP) provides daily holding period stock returns, which include dividends as well as price and cash adjustments. The *Fama-French Factors – Daily Frequency* database provides daily returns for the factors of value-weighted market index, size, book-to-market and momentum. We manually collected lists of star analysts from *Institutional Investor* magazine (October 2003 – October 2013), *The Wall Street Journal* (May 2003 – April 2013), and *StarMine* (October 2003 – August 2013). The lists of stars are matched with I/B/E/S by analysts' names and broker affiliations and double-checked for any possible inconsistencies (typos in names, analyst changes of broker in a given year, etc.). Our sample does not include analysts from some brokerage houses, notably Lehman Brothers and Merrill Lynch, as their recommendations are no longer available at I/B/E/S.

We use the following filters to the dataset. We keep only recommendations for stocks classified as ordinary shares or American Depository Receipts (CRSP Share Codes 10, 11, 12, 30, 31, and 32). To avoid the influence of "penny stocks" on our conclusions, we exclude recommendations for stocks with a price that is less than one dollar. We also exclude the recommendations from anonymous analysts or if the brokerage firm's name or code is missing.

We consider only recommendation changes and ignore re-iterations since previous research confirms that the changes carry more information than re-iterations (Boni and Womack 2006; Barber et al. 2010). Our final recommendation sample consists of three levels: "Long" (includes Strong Buy and Buy recommendations), "Hold", and "Short" (includes Sell and Strong Sell recommendations). Thus, if a particular analyst for a given company issues a Buy recommendation soon after Strong Buy, the second recommendation, that is Buy, is considered to be a re-iteration and thus it is omitted in our sample.

We use similar approach as Loh and Stulz (2011) for dealing with overall rating distribution changes that occurred primarily due to the National Association of Securities Dealers (NASD) Rule 2711 in 2002. In response to NASD Rule 2711, many brokers changed from a five-point scale to a three-point scale for their recommendations (Kadan et al. 2009). By using the I/B/E/S Stopped Recommendations File, we locate the dates when brokers stopped all previously issued recommendations, and then check the following 60 days whether a broker resumed coverage but on a three-point scale by having a new ratings distribution of either [1, 3, 5] or [2, 3, 4]. If a broker stopped the recommendations in order to re-initiate them on a three-point scale, we arrange in sequence the recommendations before and after Rule 2711 as if there were no "Stop Recommendation" signal. According to our methodology that is discussed in detail in Section 3.1 (Methods), we focus only on recommendation changes between levels. Thus, Buy and Strong Buy recommendations are considered to be on the same level, and if one follows another, we treat the latter one as re-iteration and exclude it from our analysis. The same is valid for Sell and Strong Sell recommendation. Hence, the subsequent recommendations after Rule 2711 that remain on one of the levels of Long, Hold, or Short as before Rule 2711 will be treated as re-iterations and, thus, ignored. At the same time, recommendations that were resumed on a different level (e.g. changed from Long to Hold or Short) are considered as recommendation changes and will affect our portfolios. As a result, the changes in brokers' distribution ratings do not affect the results of our tests.

Our final database contains 172,525 recommendation changes for 6443 companies listed on the NYSE, AMEX and NASDAQ markets that were announced between January 2002 and October 2014.

The entire sample of analysts is divided into the following groups:

(1) *Stars* and *Non-Stars*;
(2) *Institutional Investor* (I/I), *StarMine* Top Earnings Estimators (STM-TEE) and Top Stock Pickers (STM-TSP), and *The Wall Street Journal* (WSJ);
(3) Analysts ranked as number one (Top-Ranked): I/I-1, STM-TEE-1, STM-TSP-1 and, WSJ-1.

When a particular analyst is rated as a star in two different industries, the analyst is included only once in a particular group of stars. However, the same analyst can appear in more than one ranking group. The similarities between the lists are discussed below and are reported in Table 5.

We compare these groups using two time frames:

1) The *Year Before* is the calendar year before a ranking is announced. For example, the WSJ list of stars is announced in May 2003. Thus, the previous calendar year, from January 2002 through December 2002, is the evaluation year for the WSJ ranking. As a result, the whole sample period for *Year Before* spans from January 2002 until December 2012. We exclude the first month of January 2002 from our regression analysis because some portfolios contained too few stocks and have extraordinary returns at the beginning of that month. We evaluate analysts during the evaluation year using our portfolio approach in order to compare the rankings in a uniform way independently of the methodology used by a particular ranking.
2) The *Year After* is the one-year period that begins on the day that a particular ranking is announced and ends when the next year ranking list is announced (or twelve-month period for the last year 2013). For example, if the WSJ announcement is on May 12, 2003, the *Year After* begins on that day and ends on May 17, 2004 when the next ranking list was published. Although an entire sample period for *Year After* spans from May 2003 until October 2014, we begin by comparing groups one month after *StarMine* and I/I have published their lists, that is, from November 2003 (an incomplete month, October, is excluded from the regression analysis). We end the *Year After* period on May 2014 since that is the end of the last one-year period for the 2013 list of WSJ stars.

Table 2 displays the total number of analysts in the sample on an election-year basis. On average, approximately 13 % of analysts in our sample are listed as "stars" every year. The table shows that for every one star analyst, there are approximately six non-star analysts in our sample. Additionally, 32 % of analysts among the *Non-Stars* have been chosen as stars in some other year but not in the year under consideration. Out of 8459 analysts overall, 27 % have been elected at least once as stars, with 7 percentage points (pp) for I/I, 11 pp. – STM-TEE, 12 pp. – STM-TSP, 15 pp. – WSJ. For number-one ranked stars the figures are: I/I-1, 2.0 pp., STM-TEE-1, 4.9 pp., STM-TSP-1, 5.4 pp. and, for WSJ-1, 4.4 pp.

The average overlap among the ranking lists in each sample year is presented in Table 3. It shows the number of analysts listed by different rankings, the number of the same analysts in each pair of rankings, and the portion of the same analysts in each ranking list. Panel A presents these data for the entire groups of stars, while Panel B reports the results for the number-one ranked stars. The table also presents the percentages of analysts who appear in

Table 2 Number of analysts and the percentage of each group represented in the sample on an election-year basis. Rankings by *Institutional Investor* (I/I), Thomson Reuters' *StarMine* "Top Earnings Estimators" (STM-TEE) and "Top Stock Pickers" (STM-TSP), and *The Wall Street Journal* (WSJ). Indexation by -1 indicates a group of number-one ranked analysts. On average, 13 % of analysts were rated as stars each year. The number of star analysts in each industry varies in different rankings: for I/I it is 3 stars + Runners-up, for STM-TEE and STM-TSP – 3 stars, and for WSJ – 5 stars

Election year	All analysts	Non-stars ever elected as stars	Stars	Portion of analysts in entire groups of stars				Portion of analysts in number-one ranked stars			
				I/I	STM-TEE	STM-TSP	WSJ	I/I-1	STM-TEE-1	STM-TSP-1	WSJ-1
2003	4118	30 %	12 %	6 %	3 %	3 %	5 %	1.3 %	0.9 %	1.0 %	0.9 %
2004	4001	30 %	13 %	6 %	3 %	3 %	5 %	1.4 %	1.1 %	1.0 %	1.1 %
2005	4038	31 %	14 %	6 %	3 %	3 %	5 %	1.4 %	1.1 %	1.0 %	1.0 %
2006	4168	32 %	14 %	6 %	4 %	4 %	5 %	1.3 %	1.3 %	1.3 %	0.9 %
2007	4208	32 %	14 %	5 %	4 %	3 %	5 %	1.3 %	1.3 %	1.3 %	0.9 %
2008	4160	32 %	14 %	5 %	4 %	4 %	4 %	1.3 %	1.3 %	1.4 %	0.9 %
2009	3920	35 %	12 %	3 %	4 %	4 %	4 %	1.0 %	1.4 %	1.4 %	0.7 %
2010	3827	34 %	12 %	1 %	4 %	4 %	5 %	1.5 %	1.5 %	1.4 %	1.0 %
2011	3976	34 %	11 %	1 %	4 %	4 %	5 %	1.3 %	1.4 %	1.4 %	0.9 %
2012	3932	31 %	12 %	4 %	4 %	4 %	3 %	0.9 %	1.4 %	1.3 %	1.0 %
2013	3882	29 %	12 %	4 %	4 %	4 %	3 %	1.4 %	1.5 %	1.5 %	1.1 %
Average	4021	32 %	13 %	4 %	4 %	4 %	4 %	1.3 %	1.3 %	1.3 %	0.9 %
Overall	8459	27 %	27 %	7 %	11 %	12 %	15 %	2.0 %	4.9 %	5.4 %	4.4 %

Table 3 Average percentage of interdependence among rankings as the proportion of the same analysts in each ranking list. Panel A presents the data for entire groups of stars, while Panel B reports results for number-one ranked stars. The final line shows the average for each value. Comparisons are made on an election-year basis. Rankings by *Institutional Investor* (I/I), Thomson Reuters' *StarMine* "Top Earnings Estimators" (STM-TEE) and "Top Stock Pickers" (STM-TSP), and *The Wall Street Journal* (WSJ). Indexation by -1 indicates a group of number-one ranked analysts. The highest correlation is between WSJ and STM-TSP; the lowest is between WSJ and I/I. The number of star analysts in each industry varies in different rankings: for I/I it is 3 stars + Runners-up, for STM-TEE and STM-TSP – 3 stars, and for WSJ – 5 stars

Panel A. Entire groups of stars

Year	Proportion of the same analysts (# of the same analysts/# of star analysts)											
	WSJ in I/I	STM-TSP in I/I	STM-TEE in I/I	I/I in WSJ	STM-TSP in WSJ	STM-TEE in WSJ	I/I in STM-TSP	WSJ in STM-TSP	STM-TEE in STM-TSP	I/I in STM-TEE	WSJ in STM-TEE	STM-TSP in STM-TEE
2003	15 %	11 %	15 %	19 %	26 %	12 %	25 %	45 %	13 %	37 %	23 %	14 %
2004	9 %	8 %	12 %	12 %	28 %	10 %	18 %	47 %	11 %	26 %	17 %	10 %
2005	8 %	8 %	12 %	12 %	28 %	9 %	16 %	40 %	14 %	23 %	12 %	14 %
2006	8 %	7 %	16 %	11 %	32 %	10 %	11 %	41 %	23 %	24 %	12 %	22 %
2007	6 %	9 %	11 %	7 %	28 %	12 %	15 %	37 %	21 %	16 %	15 %	19 %
2008	6 %	10 %	11 %	8 %	26 %	10 %	15 %	32 %	22 %	16 %	11 %	21 %
2009	15 %	17 %	13 %	10 %	23 %	14 %	13 %	27 %	21 %	9 %	16 %	19 %
2010	12 %	16 %	11 %	3 %	25 %	13 %	6 %	33 %	18 %	4 %	16 %	17 %
2011	7 %	6 %	9 %	2 %	29 %	11 %	2 %	38 %	19 %	3 %	14 %	19 %
2012	4 %	7 %	10 %	7 %	25 %	7 %	8 %	20 %	22 %	10 %	5 %	20 %
2013	5 %	12 %	11 %	6 %	27 %	9 %	11 %	20 %	18 %	9 %	6 %	17 %
Avg.	9 %	10 %	12 %	9 %	27 %	11 %	13 %	35 %	18 %	16 %	13 %	18 %

Panel B. Number-one ranked analysts

Year	Proportion of the same analysts (# of the same analysts/# of star analysts)											
	WSJ-1 in I/I-1	STM-TSP-1 in I/I-1	STM-TEE-1 in I/I-1	I/I-1 in WSJ-1	STM-TSP-1 in WSJ-1	STM-TEE-1 in WSJ-1	I/I-1 in STM-TSP-1	WSJ-1 in STM-TSP-1	STM-TEE-1 in STM-TSP-1	I/I-1 in STM-TEE-1	WSJ-1 in STM-TEE-1	STM-TSP-1 in STM-TEE-1
2003	5 %	2 %	9 %	7 %	17 %	0 %	2 %	17 %	5 %	13 %	0 %	5 %
2004	2 %	2 %	7 %	2 %	14 %	2 %	2 %	14 %	2 %	9 %	2 %	2 %
2005	0 %	5 %	5 %	0 %	18 %	5 %	7 %	16 %	9 %	6 %	4 %	8 %
2006	2 %	3 %	5 %	2 %	22 %	5 %	3 %	15 %	17 %	5 %	4 %	18 %
2007	0 %	4 %	2 %	0 %	21 %	0 %	4 %	14 %	9 %	2 %	0 %	9 %
2008	0 %	5 %	5 %	0 %	15 %	5 %	5 %	10 %	10 %	5 %	4 %	11 %
2009	0 %	5 %	5 %	0 %	16 %	5 %	4 %	11 %	7 %	4 %	4 %	7 %
2010	2 %	5 %	4 %	2 %	14 %	2 %	5 %	11 %	7 %	3 %	2 %	7 %
2011	2 %	2 %	4 %	2 %	9 %	5 %	2 %	7 %	5 %	4 %	4 %	5 %

Table 3 (continued)

2012	0 %	3 %	5 %	0 %	13 %	0 %	2 %	9 %	13 %	4 %	0 %	13 %
2013	0 %	0 %	2 %	0 %	7 %	0 %	0 %	5 %	5 %	2 %	0 %	5 %
Avg.	1 %	3 %	5 %	2 %	15 %	3 %	3 %	12 %	8 %	5 %	2 %	8 %

another ranking. For example, the *Institutional Investor* ranking has, on average, 13 % of analysts out of 191 unique names who were listed as "Top Stock Pickers" by *StarMine* in the same years. As can be observed, Top Stock Pickers and *The Wall Street Journal* exhibit the highest similarity in their published lists, while *Institutional Investor* and *The Wall Street Journal* have the lowest similarity. Such overlap is expected given the degree of similarities in the evaluation methods used. It also shows how different the lists of Star analysts are, which might explain the differences in the returns from their recommendations.

Table 4 shows the number of firms in the sample, which ranged from 3356 in 2010 to 3935 in 2006, and the percentage of firms covered by each group. On average per year, I/I, STM-TEE, and STM-TSP star analysts covered 22–23 %, while WSJ covers 27 % of the firms in the sample. Out of the total number of 6443 firms in our sample, I/I stars covered 36 % of the firms, STM-TEE – 47 %, STM-TSP – 50 %, and WSJ – 58 %. This difference suggests that these groups have different firm coverage, which could be explained by the fact that the WSJ list has the highest turnover of names (they issue recommendations for different universes of firms). Number-one ranked stars cover 10 % of the firms, except WSJ-1, which covers 8 % of the firms.

As seen in Table 5, the group of *Stars* issues on average 20 % of all recommendations in our sample. Both WSJ and I/I stars issue more recommendations than STM-TEE and STM-TSP stars.

Table 4 Number of firms and percentage of firms in the sample covered by each group, calculated on an election-year basis. Rankings by *Institutional Investor* (I/I), Thomson Reuters' *StarMine* "Top Earnings Estimators" (STM-TEE) and "Top Stock Pickers" (STM-TSP), and *The Wall Street Journal* (WSJ). Indexation by -1 indicates a group of number-one ranked analysts. Each group of star analysts covers approximately 50 % of firms in the sample. Thus, the coverage universe differs for the various groups of stars. The number of star analysts in each industry varies in different rankings: for I/I it is 3 stars + Runners-up, for STM-TEE and STM-TSP – 3 stars, and for WSJ – 5 stars

Election year	Total number of firms	Portion of firms covered by								
		Stars	Entire groups of stars				Number-one ranked stars			
			I/I	STM-TEE	STM-TSP	WSJ	I/I-1	STM-TEE-1	STM-TSP-1	WSJ-1
2003	3580	50 %	31 %	18 %	19 %	29 %	12 %	7 %	7 %	9 %
2004	3660	50 %	27 %	17 %	20 %	29 %	10 %	8 %	8 %	9 %
2005	3801	50 %	27 %	19 %	19 %	29 %	9 %	8 %	8 %	8 %
2006	3935	48 %	24 %	22 %	22 %	26 %	9 %	9 %	10 %	7 %
2007	3927	53 %	24 %	22 %	22 %	30 %	9 %	9 %	10 %	7 %
2008	3752	51 %	25 %	25 %	25 %	27 %	10 %	13 %	13 %	8 %
2009	3495	49 %	18 %	25 %	25 %	28 %	9 %	10 %	11 %	8 %
2010	3356	52 %	12 %	27 %	27 %	29 %	12 %	12 %	12 %	8 %
2011	3414	49 %	12 %	23 %	22 %	29 %	12 %	10 %	9 %	8 %
2012	3390	49 %	21 %	25 %	22 %	20 %	9 %	11 %	11 %	9 %
2013	3425	51 %	24 %	24 %	25 %	20 %	10 %	12 %	12 %	8 %
Average	3612	50 %	22 %	22 %	23 %	27 %	10 %	10 %	10 %	8 %
Overall	6443	71 %	36 %	47 %	50 %	58 %	20 %	29 %	32 %	30 %

Table 5 Number of recommendations and the percentage of recommendations on an election-year basis by each ranking. Rankings by *Institutional Investor* (I/I), Thomson Reuters' *StarMine* "Top Earnings Estimators" (STM-TEE) and "Top Stock Pickers" (STM-TSP), and *The Wall Street Journal* (WSJ). Indexation by -1 indicates a group of number-one ranked analysts. The number of star analysts in each industry varies in different rankings: for I/I it is 3 stars + Runners-up, for STM-TEE and STM-TSP – 3 stars, and for WSJ – 5 stars

Election year	Entire sample	*Stars*	Entire groups of stars				Number-one ranked stars			
			I/I	STM-TEE	STM-TSP	WSJ	I/I-1	STM-TEE-1	STM-TSP-1	WSJ-1
2003	33,233	20 %	10 %	4 %	5 %	8 %	2.0 %	1.3 %	1.3 %	1.7 %
2004	32,600	20 %	9 %	4 %	5 %	8 %	1.9 %	1.5 %	1.6 %	1.9 %
2005	32,175	21 %	9 %	4 %	5 %	8 %	1.5 %	1.6 %	1.5 %	1.8 %
2006	32,005	20 %	8 %	6 %	6 %	7 %	1.7 %	2.1 %	2.2 %	1.4 %
2007	34,254	27 %	8 %	5 %	10 %	12 %	1.8 %	1.6 %	2.0 %	5.9 %
2008	36,551	25 %	7 %	6 %	7 %	11 %	1.8 %	2.7 %	2.9 %	6.0 %
2009	33,217	22 %	5 %	5 %	10 %	11 %	1.5 %	1.6 %	2.1 %	4.9 %
2010	31,686	17 %	2 %	6 %	6 %	7 %	2.1 %	2.3 %	2.3 %	1.3 %
2011	30,607	16 %	2 %	5 %	4 %	7 %	2.0 %	1.8 %	1.6 %	1.7 %
2012	29,181	18 %	6 %	6 %	5 %	4 %	1.6 %	2.0 %	2.3 %	1.7 %
2013	27,412	19 %	6 %	6 %	6 %	5 %	1.8 %	2.4 %	2.2 %	1.6 %
Average	32,084	20 %	6 %	5 %	6 %	8 %	1.8 %	1.9 %	2.0 %	2.7 %
Overall	291,731	19 %	6 %	6 %	7 %	9 %	1.7 %	2.1 %	2.3 %	2.7 %

3 Results: risk-adjusted portfolio returns

3.1 Methods

To measure the profitability of the recommendations, we apply a well-established methodology by constructing dynamic portfolios. We construct buy-and-hold "Long", "Hold", and "Short" portfolios for each sub-group of analysts in the year subsequent to the year in which the rankings were assigned (referred to as *Year After*) and for the year during which the analysts were evaluated (referred to as *Year Before*) (Barber et al. 2006; Fang and Yasuda 2014). For each new Strong Buy or Buy recommendation, $1 is invested at the end of the recommendation announcement day (or at the close of the next trading day if the recommendation is issued after the closing of trading or on a non-trading day) into the "Long" portfolio. The stock is held in the portfolio for the following calendar year if there are no recommendation revisions or recommendation changes by the same analyst. If, during the following year, the analyst changes his or her recommendation level from Strong Buy or Buy to Hold or Sell or Strong Sell, then the stock is withdrawn from the "Long" portfolio and placed in the "Hold" or "Short" portfolio by the end of the trading day on which the new recommendation is issued (or at the close of the next trading day if the recommendation is issued after the closing of trading or on a non-trading day). If there is a recommendation revision, but the new recommendation is on the same level (that is, Buy or Strong Buy), then the stock is not kept in the same portfolio for an additional calendar year, but only until the next recommendation change within one year from the initial recommendation. Thus, re-

iterations of recommendations are not included in the portfolios. The same procedures are applied to a "Hold" (includes only Hold recommendations) and "Short" (includes Sell and Strong Sell recommendations) portfolios. As a result of this strategy, the calendar day t gross return on portfolio ρ includes from $n = 1$ to $N_{\rho t}$ recommendations and could be defined as:

$$R_{\rho t} = \frac{\sum_{n=1}^{N_{\rho t}} X_{n,t-1} R_{i_n,t}}{\sum_{n=1}^{N_{\rho t}} X_{n,t-1}}, \tag{1}$$

where $X_{n,\ t\text{-}1}$ is the cumulative total gross return of stock i_n from the next trading day after a recommendation was added to the portfolio to day t-1, which is the previous trading day before t, that is:

$$X_{n,t-1} = R_{i_n,\mathrm{recdat}_n+1} R_{i_n,\mathrm{recdat}_n+2} * \ldots * R_{i_n,\mathrm{recdat}_n t-1} \tag{2}$$

Daily excess returns for each group's "Long", "Hold" and "Short" portfolios are estimated as an intercept (alpha) that is calculated according to the four-factor model proposed by (Carhart 1997):

$$R_{\rho\tau} - Rf_\tau = \alpha_\rho + \beta_\rho (Rm_\tau - Rf_\tau) + s_\rho SMB_\tau + h_\rho HML_\tau + m_\rho UMD_\tau + \varepsilon_{\rho\tau}, \tag{3}$$

where

- Rm_τ is a daily market return
- Rf_τ is the risk-free rate of return, SMB_τ is a size factor, that is, the difference between the value-weighted portfolio returns of small and large stocks
- HML_τ is a book-to-market factor, that is, the difference between the value-weighted portfolio returns of high book-to-market and low book-to-market stocks
- UMD_τ is a momentum factor, that is, the difference in the returns of stocks with a positive return momentum and those with a negative return momentum over months τ-12 and τ-2.

The alpha differentials (differences in alphas) are statistically tested using two approaches. Alphas for groups in the same year, that is, *Year After* or *Year Before*, are compared using daily differences in gross returns, which are regressed on four factors according to Equation (3). An intercept from this regression returns the difference in alpha, and a t-test indicates whether this difference is statistically significant. To compare excess returns between *Year After* with *Year Before*, the seemingly unrelated estimation is accompanied by a test for significant differences in the intercepts from various regressions (*suest* and *test* procedures in STATA).[4]

Even though all reported excess returns and alpha differentials are calculated on a daily basis, we report figures in monthly values by multiplying daily values with 21 trading days.

[4] We repeat our analysis using SUEST test for the statistical significance of alpha differentials for contemporaneous comparison of the groups in the *Year After* or *Year Before*. Qualitatively, our results remain the same as those reported in Tables 9 and 10.

3.2 Results and discussion

Table 6 represents the average monthly excess returns (alphas) for "*Stars*" and "*Non-Stars*" during the year after rankings have been published (Panel A), during the evaluation year (Panel B) and as a comparison of the returns in the *Year After* with those of the *Year Before* (Panel C). The first three rows in each panel of the table show the returns of the Long, Hold and Short

Table 6 Average monthly abnormal returns (alphas) for groups of *Star* and *Non-Star* analysts and differences in abnormal returns. Rankings by *Institutional Investor* (I/I), Thomson Reuters' *StarMine* "Top Earnings Estimators" (STM-TEE) and "Top Stock Pickers" (STM-TSP), and *The Wall Street Journal* (WSJ). Portfolios are built according to recommendations: when a new recommendation is announced, $1 is invested in the recommended stock by the end of the trading day (or on the next trading day if the recommendation is issued after the close of trading or is announced on a non-trading day), and the stock is held for one year or until the same analyst changes his or her recommendation or drops coverage, in which case the stock is withdrawn by the end of that trading day. All figures are obtained as intercepts from the regressions of the daily returns time series from two sample periods: the *Year Before* (February 2002 – December 2012) and the *Year After* (November 2003 – May 2014) on four standard risk factors (Carhart's four-factor model). The Long portfolio includes Buy and Strong Buy recommendations; Hold portfolio includes all Hold recommendations, while the Short portfolio includes Sell, and Strong Sell recommendations. Portfolios by *Stars* exhibit statistically significant decrease in performance from the *Year Before* to the *Year After*. Long-Short portfolio by *Stars* performs insignificantly from that of the *Non-Stars*. Only Buy and Strong Buy Recommendations by Star analysts persistently outperform *Non-Stars*

Portfolio	Average monthly abnormal return (%) for the overall groups		
	Stars	Non-Stars	Difference Stars – Non-Stars
Panel A. Year After (November 2003 – May 2014)			
Long: Strong Buy/Buy	0.33***	0.18*	0.15**
	(4.09)	(1.75)	(2.03)
Hold	0.03	−0.07	0.10
	(0.45)	(−0.73)	(1.57)
Short: Sell/ Strong Sell	−0.20	−0.29**	0.09
	(−1.56)	(−2.42)	(0.78)
Long-Short	0.53***	0.47***	0.06
	(4.56)	(5.46)	(0.56)
Panel B. Year Before (February 2002 – December 2012)			
Long	0.62***	0.16	0.45***
	(7.28)	(1.46)	(5.14)
Hold	0.01	−0.08	0.10
	(0.15)	(−0.80)	(1.24)
Short	−0.46***	−0.24	−0.22*
	(−3.44)	(−1.62)	(−1.67)
Long-Short	1.08***	0.40***	0.68***
	(9.00)	(3.71)	(5.30)
Panel C. Difference Year After – Year Before (SUEST TEST)			
Long	−0.29***	0.02	–
Hold	0.02	0.01	–
Short	0.26**	−0.05	–
Long-Short	−0.55***	0.07	–

t-statistics in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

portfolios, while the third row (Long-Short) presents the total return on all of the recommendations for a particular group, which is the Long minus the Short portfolio returns.

As we can see in Table 6, Panel A, the Long-Short portfolio of *Stars*, with monthly alphas of +0.53 %, performed insignificantly different from the *Non-Stars*, with monthly alphas of +0.47 %, leading to a statistically insignificant difference of 0.06 percentage points in abnormal returns for a Long-Short portfolio in the year after rankings were published. For alpha differentials among all portfolios in the *Year After*, only Long portfolio of *Stars* is statistically different from that of the *Non-Stars*. As can be expected, during the evaluation year (Panel B in Table 6), *Stars* had higher recommendation returns, of +1.08 %, than *Non-Stars*, with +0.40 %. When we analyze the differences in returns from the *Year Before* to the *Year After*, as reported in Panel C, we conclude that the *Stars* do not continue to perform on the same level, which is reflected as a significant difference in the returns on their Long, Short and Long-Short portfolios, while the group of *Non-Stars* had an insignificant difference in the returns on their Long, Short and Long-Short portfolios. Hence, we conclude that only Buy and Strong Buy recommendations of the group of *Stars* persistently outperform those of their non-star peers, although the *Stars* show a decrease in their performance in the *Year After*.

Table 7 shows the excess returns from recommendations issued by entire groups of stars: "I/I", "STM-TEE", "STM-TSP" and "WSJ" in the *Year After* (Panel A) and *Year Before* (Panel B) and the difference in returns between the *Year After* and *Year Before* (Panel C).

Long-Short portfolios from the groups of STM-TEE, STM-TSP and WSJ stars show positive and statistically significant alphas in both time periods, while I/I stars performed insignificantly different from the market in the *Year After*. Excess returns of the Long portfolios for all groups in the *Year Before* and *Year After* are statistically different from zero. Some Short portfolios are insignificantly different from zero (that of I/I, STM-TSP in the *Year After*, and I/I, STM-TEE in the *Year Before*).

As can be seen in Panel C of Table 7, Long, Short and Long-Short portfolios from WSJ and STM-TSP stars exhibit the greatest significant decrease in performance after election as a star, that is −1.48 % for WSJ, and −1.37 % for STM-TSP for their Long-Short portfolios. This decrease can be explained as the regression to the mean, which shows that it is very difficult to issue recommendations consistently generating portfolios with very high abnormal returns. At the same time, STM-TEE shows persistence because their Long-Short portfolio in the *Year Before* is insignificantly different to the *Year After* (albeit the group of STM-TEE has significantly different performance for their Short portfolio, as we can see in Panel C). For I/I stars, the drop is −0.26 %.

Table 8 shows the average monthly excess returns for the top-ranked analysts (number-one ranked analysts) for the *Year After* election (Panel A) and the *Year Before* (Panel B) and the difference between the *Year After* and the *Year Before* (Panel C). We find that the Hold portfolios for all groups and all time periods perform insignificantly different from the market, having alphas insignificantly different from zero.

In the *Year After* election, the groups of STM-TEE-1 and I/I-1 stars show positive and statistically significant alphas for their Long-Short portfolios, while STM-TSP-1 and WSJ-1 stars performed on the market level, which is explained by relatively low and insignificant alphas for their Short portfolios. The highest return, and the only statistically significant one among the Short portfolios in the *Year After*, was generated by the STM-TEE-1 group, with −0.70 %. Excess returns of Long portfolios for all groups in the *Year After* are positive and significantly different from zero. Panel C shows that Long, Short and Long-Short portfolios from STM-TSP-1 and WSJ-1 stars show a statistically significant decrease in performance from the *Year Before* to the *Year After*, while I/I-1 and STM-TEE-1 exhibit persistence from the *Year Before* to the *Year After*. Hence, comparing the returns in the *Year After* election with the *Year Before* election in Panel C of Table 8,

Table 7 Average monthly abnormal returns (alphas) for each group of star analysts. Rankings by *Institutional Investor* (I/I), Thomson Reuters' *StarMine* "Top Earnings Estimators" (STM-TEE) and "Top Stock Pickers" (STM-TSP), and *The Wall Street Journal* (WSJ). Portfolios are built according to recommendations: when a new recommendation is announced, $1 is invested in the recommended stock by the end of the trading day (or on the next trading day if the recommendation is issued after the closing of trading or is announced on a non-trading day), and the stock is held for one year or until the same analyst changes his or her recommendation or drops coverage, in which case the stock is withdrawn by the end of that trading day. All figures are obtained as intercepts from the regressions of the daily returns time series on four standard risk factors (Carhart's four-factor model) and represented in monthly values from two sample periods: the *Year Before* (February 2002 – December 2012) and the *Year After* (November 2003 – May 2014). The Long portfolio includes Buy and Strong Buy recommendations; Hold portfolio includes all Hold recommendations; the Short portfolio includes Sell and Strong Sell recommendations. Groups of STM-TEE, STM-TSP and WSJ stars show positive and statistically significant alphas for their Long-Short portfolios in both time periods, while recommendations from I/I stars performed on the market level. Excess returns of Long portfolios for all groups in the *Year Before* and *Year After* are significantly different from zero. Some Short portfolios are insignificantly different from zero (that of I/I, STM-TSP in the *Year After*, and I/I in the *Year Before*). Long, Short and Long-Short portfolios from STM-TSP and WSJ stars show significant decrease in the performance from the *Year Before* to the *Year After*. STM-TEE shows persistency from the *Year Before* to the *Year After* (all differences are insignificant in Panel C). Hold portfolios for all groups and all time periods perform on the market level (except of I/I in the *Year After*)

Portfolio	Average monthly abnormal returns (%) for the entire groups of star analysts			
	I/I	STM-TEE	STM-TSP	WSJ
Panel A. Year After (November 2003 – May 2014)				
Long: Strong Buy/Buy	0.32***	0.32***	0.38***	0.23***
	(3.33)	(3.09)	(3.83)	(2.18)
Hold	0.16*	0.01	-0.02	-0.02
	(1.70)	(0.10)	(−0.24)	(−0.21)
Short: Sell/ Strong Sell	0.18	-0.64***	-0.16	-0.40**
	(1.12)	(−3.44)	(−0.79)	(−2.02)
Long-Short	0.14	0.97***	0.54***	0.63***
	(0.96)	(5.19)	(2.69)	(3.10)
Panel B. Year Before (February 2002 – December 2012)				
Long	0.39***	0.36***	0.90***	0.98***
	(3.70)	(3.20)	(8.13)	(8.27)
Hold	0.05	0.05	0.05	0.01
	(0.48)	(0.45)	(0.44)	(0.11)
Short	-0.01	-0.24	-1.02***	-1.14***
	(−0.07)	(−1.08)	(−5.10)	(−5.66)
Long-Short	0.40**	0.59***	1.91***	2.11***
	(2.37)	(2.68)	(9.58)	(10.84)
Panel C. Difference Year After – Year Before (SUEST TEST)				
Long	-0.07	-0.04	-0.52***	-0.75***
Hold	0.11	-0.04	-0.07	-0.03
Short	0.19	-0.40*	0.86***	0.74***
Long-Short	-0.26*	0.38	-1.37***	-1.48***

t-statistics in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

the returns of STM-TSP-1 and WSJ-1 decrease, with a significant difference of −1.61 and −2.67 % between alphas in the *Year After* and *Year Before*, respectively.

Table 8 Average monthly abnormal returns (alphas) for each group of number-one ranked analysts. Rankings by *Institutional Investor* (I/I), Thomson Reuters' *StarMine* "Top Earnings Estimators" (STM-TEE) and "Top Stock Pickers" (STM-TSP), and *The Wall Street Journal* (WSJ). Indexation by -1 signifies a group of number-one ranked analysts. Portfolios are built according to recommendations: when a new recommendation is announced, $1 is invested in the recommended stock by the end of the trading day (or on the next trading day if the recommendation is issued after the closing of trading or is announced on a non-trading day), and the stock is held for one year or until the same analyst changes his or her recommendation or drops coverage, in which case the stock is withdrawn by the end of that trading day. All figures are obtained as intercepts from the regressions of the daily returns time series on four standard risk factors (Carhart's four-factor model) and represented in monthly values from two sample periods: the *Year Before* (February 2002 – December 2012) and the *Year After* (November 2003 – May 2014). The Long portfolio includes Buy and Strong Buy recommendations; Hold portfolio includes all Hold recommendations; the Short portfolio includes Sell, and Strong Sell recommendations. Groups of STM-TEE-1 and I/I-1 stars show positive and statistically significant alphas for their Long-Short portfolios in the *Year After*. Long-Short portfolios from STM-TSP-1 and WSJ-1 stars performed on the market level which is explained by low and insignificant alphas for their Short portfolios. Excess returns of Long portfolios for all groups in the *Year After* are positive and significantly different from zero. Long, Short and Long-Short portfolios from STM-TSP-1 and WSJ-1 stars show significant decrease in the performance from the *Year Before* to the *Year After*, while I/I-1 and STM-TEE-1 exhibit the persistently from the *Year Before* to the *Year After* (all differences are insignificant in Panel C). Hold portfolios for all groups and all time periods perform on the market level

Portfolio	Average monthly abnormal returns (%) for number-one ranked star analysts			
	I/I-1	STM-TEE-1	STM-TSP-1	WSJ-1
Panel A. Year After (November 2003 – May 2014)				
Long: Strong Buy/Buy	0.44***	0.31**	0.42***	0.38**
	(3.26)	(2.20)	(3.31)	(2.11)
Hold	0.09	-0.17	0.02	-0.02
	(0.66)	(−1.31)	(0.17)	(−0.16)
Short: Sell/ Strong Sell	-0.27	-0.70**	0.23	-0.20
	(−1.14)	(−2.44)	(0.59)	(−0.46)
Long-Short	0.71***	1.01***	0.18	0.58
	(2.83)	(3.34)	(0.46)	(1.26)
Panel B. Year Before (February 2002 – December 2012)				
Long	0.51***	0.31*	0.93***	1.35***
	(3.45)	(1.93)	(5.83)	(6.98)
Hold	0.12	0.19	-0.12	0.02
	(0.81)	(1.28)	(−0.82)	(0.11)
Short	-0.35	-0.28	-0.85***	-1.90***
	(−1.34)	(−0.84)	(−2.78)	(−4.96)
Long-Short	0.87***	0.58*	1.79***	3.25***
	(2.98)	(1.66)	(5.34)	(7.97)
Panel C. Difference Year After – Year Before (SUEST TEST)				
Long	-0.07	0.00	-0.51***	-0.97***
Hold	-0.03	-0.36**	0.14	-0.04
Short	0.08	-0.42	1.08**	1.70***
Long-Short	-0.16	0.43	-1.61***	-2.67***

t-statistics in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In Tables 9 and 10, we report the alpha differentials obtained by comparing the abnormal returns between groups of stars in the *Year Before* and *Year After*. In both tables, we report

results for Long in Panel A, Short – Panel B, and Long-Short – Panel C. First, we discuss whether each particular group of stars outperformed the *Non-Stars*. Then, we comment on the differences in performance among all of the groups of star analysts.

Comparing the returns of the Long portfolios of all of the groups of stars with those of *Non-Stars* in the *Year Before* (Panel A in Table 9), we find that the groups of I/I, STM-TSP and WSJ stars and their number-one sub-groups significantly outperformed *Non-Stars*, while returns of the STM-TEE and number-one ranked STM-TEE-1 had insignificant differences with *Non-Stars*. Similar results were observed for the Short portfolios (Panel B in Table 9), where only STM-TSP and WSJ as well as their number-one sub-groups significantly outperformed *Non-Stars*. When we combine the performance of the Long and Short portfolios, that is the Long-Short portfolio (Panel C), we can see that STM-TSP and WSJ stars outperformed the *Non-Stars*. Such a finding is expected considering the similarity of the methods used by STM-TSP and WSJ rankings (recommendation-based) and our performance measures.

In line with the fact that STM-TSP and WSJ are based on the investment value of recommendations, and the methodologies of I/I and STM-TEE do not focus on the performance of the recommendations, we find that the Long, Short, Long-Short portfolios from the groups of STM-TSP and WSJ and their sub-groups of the number-one stars perform significantly better than both groups of I/I and STM-TEE and their number-one stars I/I-1 and STM-TEE-1 (Panel A, B and C in Table 9). Considering the returns for the Long and Short portfolios in the *Year Before*, the Long-Short portfolios represent similar patterns: stars ranked by the methodologies that take into account investment value of recommendations (STM-TSP and WSJ) outperformed those that do not use the recommendations in their election methods (I/I and STM-TEE), with those differences being statistically significant. This result is expected given the evaluation methodologies applied to rank stars, since both STM-TSP and WSJ rankings are objectively measuring the investment value of the recommendations, whereas STM-TEE focuses on the accuracy and timing of earnings forecasts, and the I/I ranking uses investment value as one of several attributes in its subjective methodology (survey) to select stars.

Comparing the returns of the Long portfolios in the *Year After* of all of the groups of stars with those of *Non-Stars* (first column in Panel A of Table 10), we find that only returns of I/I-1, STM-TSP and STM-TSP-1 stars significantly outperformed those of the group of *Non-Stars* at the 10 % significance level, while entire groups of I/I, STM-TEE, and WSJ as well as number-one ranked stars from STM-TEE-1 and WSJ-1 did not significantly differ from *Non-Stars*. Analyzing Panel B of Table 10 for the Short portfolios in the *Year After*, we find that the differences between the excess returns of all of the groups of stars and those of *Non-Stars* were insignificant, except for I/I who underperformed, and STM-TEE who outperformed, the *Non-Stars*. This high performance of the Short portfolio for the STM-TEE and low performance for I/I was reflected in how the returns of these groups differ from the others. Hence, the differences in returns among most of the Short portfolios in the *Year After* are insignificant, except for I/I being significantly lower than some other groups of stars, and STM-TEE being significantly better than STM-TSP and their sub-group of STM-TSP-1. The results for the Long and Short portfolios explain why Long-Short portfolios of I/I underperformed *Non-Stars* and some of the groups of stars, while the STM-TEE and their number-one analysts performed significantly better than the *Non-Stars* and STM-TSP. We interpret this result to reflect the importance of making accurate earnings forecasts[5] in order to predict a future price decrease,

[5] As has been mentioned previously, earnings forecasts are important for valuation of stocks and thus for predicting future stock price (Ohlson 1995; Loh and Mian 2006).

Table 9 Alpha differentials in the *Year Before* calculated as the difference in the excess return from the vertical group minus the excess return for a horizontal group of stars. Rankings by *Institutional Investor* (I/I), Thomson Reuters' *StarMine* "Top Earnings Estimators" (STM-TEE) and "Top Stock Pickers" (STM-TSP), and *The Wall Street Journal* (WSJ). Indexation by -1 signifies a group of number-one ranked analysts. Excess returns were obtained from regressions for time series from a sample period of the *Year Before* (February 2002 – December 2012). Long portfolios (Panel A) include Strong Buy and Buy recommendations; Short portfolios (Panel B) include Sell and Strong Sell recommendations. Panel C shows the results for the Long-Short portfolios. Hold portfolios are not compared

	Alpha differentials (%) in the Year Before (February 2002 – December 2012) for the groups of analysts							
	Non-Stars	I/I	STM-TEE	STM-TSP	WSJ	I/I-1	STM-TEE-1	STM-TSP-1
Panel A. Long portfolios: Strong Buy, Buy recommendations								
I/I	0.23**							
STM-TEE	0.19	-0.03						
STM-TSP	0.73***	0.51***	0.54***					
WSJ	0.81***	0.58***	0.62***	0.08				
I/I-1	0.35**	0.12	0.15	-0.39***	-0.46***			
STM-TEE-1	0.14	-0.08	-0.05	-0.59***	-0.67***	-0.20		
STM-TSP-1	0.77***	0.54***	0.57***	0.03	0.04	0.42**	0.62***	
WSJ-1	1.19***	0.96***	0.99***	0.45***	0.38***	0.84***	1.04***	0.42**
Panel B. Short portfolios: Sell, Strong Sell recommendations								
I/I	0.23							
STM-TEE	0.00	-0.22						
STM-TSP	-0.78***	-1.00***	-0.78***					
WSJ	-0.90***	-1.13***	-0.90***	-0.12				
I/I-1	-0.12	-0.34*	-0.12	0.66**	0.78***			
STM-TEE-1	-0.04	-0.26	-0.04	0.74**	0.86**	0.08		
STM-TSP-1	-0.62**	-0.84***	-0.62*	0.16	0.28	-0.50	-0.58	
WSJ-1	-1.66***	-1.88***	-1.66***	-0.88**	-0.76**	-1.54***	-1.62***	-1.04**
Panel C. Long-Short portfolios								
I/I	0.00							
STM-TEE	0.19	0.19						
STM-TSP	1.51***	1.51***	1.32***					
WSJ	1.71***	1.71***	1.52***	0.20				
I/I-1	0.47	0.46**	0.27	-1.05***	-1.25***			
STM-TEE-1	0.18	0.18	-0.01	-1.33***	-1.53***	-0.28		
STM-TSP-1	1.38***	1.38***	1.19***	-0.13	-0.33	0.92**	1.20***	
WSJ-1	2.85***	2.84***	2.65***	1.33***	1.13***	2.38***	2.66***	1.46***

*** p < 0.01. ** p < 0.05. * p < 0.1

which facilitated the outperformance of the Short portfolios of the STM-TEE stars in order to outperform *Non-Stars* as well as STM-TSP stars.

By analyzing the persistence in the performance for the Long portfolios from the *Year Before* to the *Year After*, we conclude that the groups of I/I-1, STM-TSP and STM-TSP-1 performed significantly better than *Non-Stars* in both the *Year After* and *Year Before* (Panel A in Tables 9 and 10), while the group of STM-TEE with their sub-group of number-one ranked

Table 10 Alpha differentials in the *Year After* calculated as the difference in the excess return from the vertical group minus the excess return for a horizontal group of stars. Rankings by *Institutional Investor* (I/I), Thomson Reuters' *StarMine* "Top Earnings Estimators" (STM-TEE) and "Top Stock Pickers" (STM-TSP), and *The Wall Street Journal* (WSJ). Indexation by -1 signifies a group of number-one ranked analysts. Excess returns were obtained from regressions for time series from a sample period of the *Year After* (November 2003 – May 2014). Long portfolios (Panel A) include Strong Buy and Buy recommendations; Short portfolios (Panel B) include Sell and Strong Sell recommendations. Panel C shows the results for the Long-Short portfolios. Hold portfolios are not compared

	Alpha differentials (%) in the Year After (November 2003 – May 2014) for the groups of analysts							
	Non-Stars	I/I	STM-TEE	STM-TSP	WSJ	I/I-1	STM-TEE-1	STM-TSP-1
Panel A. Long portfolios: Strong Buy, Buy recommendations								
I/I	0.15							
STM-TEE	0.15	0.00						
STM-TSP	0.21*	0.06	0.06					
WSJ	0.05	-0.09	-0.09	-0.15				
I/I-1	0.26*	0.12	0.12	0.06	0.21			
STM-TEE-1	0.14	-0.01	-0.01	-0.07	0.08	-0.13		
STM-TSP-1	0.24*	0.09	0.09	0.03	0.19	-0.03	0.10	
WSJ-1	0.21	0.06	0.06	0.00	0.15	-0.06	0.07	-0.03
Panel B. Short portfolios: Sell, Strong Sell recommendations								
I/I	0.47***							
STM-TEE	-0.35**	-0.82***						
STM-TSP	0.13	-0.34	0.48**					
WSJ	-0.11	-0.58***	0.24	-0.24				
I/I-1	0.02	-0.45**	0.37	-0.11	0.13			
STM-TEE-1	-0.41	-0.88***	-0.05	-0.53*	-0.29	-0.43		
STM-TSP-1	0.52	0.05	0.87**	0.39	0.63	0.50	0.93**	
WSJ-1	0.09	-0.38	0.44	-0.04	0.20	0.07	0.49	-0.43
Panel C. Long-Short portfolios								
I/I	-0.32**							
STM-TEE	0.50***	0.82***						
STM-TSP	0.08	0.40*	-0.42*					
WSJ	0.16	0.49**	-0.34	0.09				
I/I-1	0.24	0.57***	-0.26	0.17	0.08			
STM-TEE-1	0.54*	0.86***	0.04	0.47	0.38	0.30		
STM-TSP-1	-0.28	0.04	-0.78*	-0.36	-0.44	-0.53	-0.82*	
WSJ-1	0.12	0.44	-0.38	0.04	-0.05	-0.13	-0.42	0.40

*** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$

STM-TEE-1 stars had insignificantly different returns from *Non-Stars* before and after rankings were announced. Additionally, it is important to mention that the difference in the returns of the Long portfolios in the *Year Before* (Panel A in Table 9) among the entire groups of star analysts shows that, while STM-TSP and WSJ stars and their sub-groups of number-one STM-TSP-1 and WSJ-1 stars significantly outperformed I/I, STM-TEE and their number-one sub-groups, the differences in returns among all of the groups in the *Year After* are insignificant

(Panel A in Table 10). This result confirms the assumption that, in most cases, there is the regression to the mean, which explains why in the *Year Before* the differences in returns were mostly statistically significant while in the *Year After* all of the groups of stars perform insignificantly different from each other.

Figure 1 shows a comparison of frequency of months when a particular group appears to be the best group compared with other groups within the same comparison pool. For example, using raw monthly returns (calculated as geometrically-aggregated daily returns within a given month), *Stars* are compared with *Non-Stars*: the number of months when *Stars* outperformed *Non-Stars* is divided by the total number of months in the sample period. These results are in line with the abnormal returns analyzed above. We observe that Long portfolio of Stars outperformed *Non-Stars* for 59 % of the months in our sample period. However, the Short portfolio of *Stars* outperformed that of the *Non-Stars* during only 46 % of the months. In the pool with the entire groups of stars for both Long and Short portfolios, STM-TEE, STM-TSP and WSJ stars show very similar frequencies of months when a given group outperformed the others, and only I/I stars had a substantially lower number of months than the other three groups. For top-ranked analysts, the Long and Short portfolios of all sub-groups of stars exhibit similar patterns with the WSJ-1 analysts having the highest frequency for both portfolios.

Using the returns obtained by comparing the portfolio returns in the years after election and analyzing the frequency of months in which particular groups outperformed the others, we conclude as follows: Firstly, we document that only the Long portfolio of the group of *Stars* which has all unique names of ranked analysts from I/I, WSJ and *StarMine* rankings performs better than that of the *Non-Stars*. Note, though, that Short and Long-Short (Long minus Short) portfolios from both groups have statistically insignificant differences in returns. Our finding that the group of *Stars* outperforms *Non-Stars* only for Buy and Strong Buy but not for Sell and Strong Sell recommendations is consistent with a hypothesis that reputation helps in mitigating conflicts of interest. According to Fang and Yasuda (2009), reputation effects from star rankings can improve the quality of analyst research and this improvement might happen more for favorable ratings like Buy and Strong Buy, which are more likely to be conflicted (Michaely and Womack 1999; Barber et al. 2006).

Considering that, in the year before election, *Stars* had significantly higher returns for their Long-Short portfolios and the differences between Long and Short portfolios were greater than in the year after, we conclude that, on average, *Stars* are not able to repeat the observed high excess returns after they were selected as *Stars*. However, this result does not imply that all rankings' methodologies have the same low predictive power of the future returns.

Secondly, we investigated the differences among subjective (I/I) and objective (*StarMine* and WSJ) rankings and found that the returns of the recommendations from analysts ranked by the subjective ranking from the *Institutional Investor* had relatively low returns in the *Year Before* as well as in the *Year After* election. Indeed, most of the portfolios from I/I stars underperformed the other groups of stars from objective rankings and even underperformed the non-star peers. Though this conclusion might be essential for the investors who are concerned about the investment value of recommendations, it is important to remember that the *Institutional Investor* ranking does not consider the recommendations as the primary evaluation criteria for selecting analysts. According to a list of attributes mentioned by the *Institutional Investor* magazine and to Table 1 in Bagnoli et al. (2008), stock selection is not the number one criteria institutions are looking for when ranking analysts. In fact, stock selection is typically not even among the top five attributes that institutional investors are looking for. Thus, our findings do not contradict the intentions of the I/I ranking (that is, to evaluate subjectively the services provided by sell-side analysts).

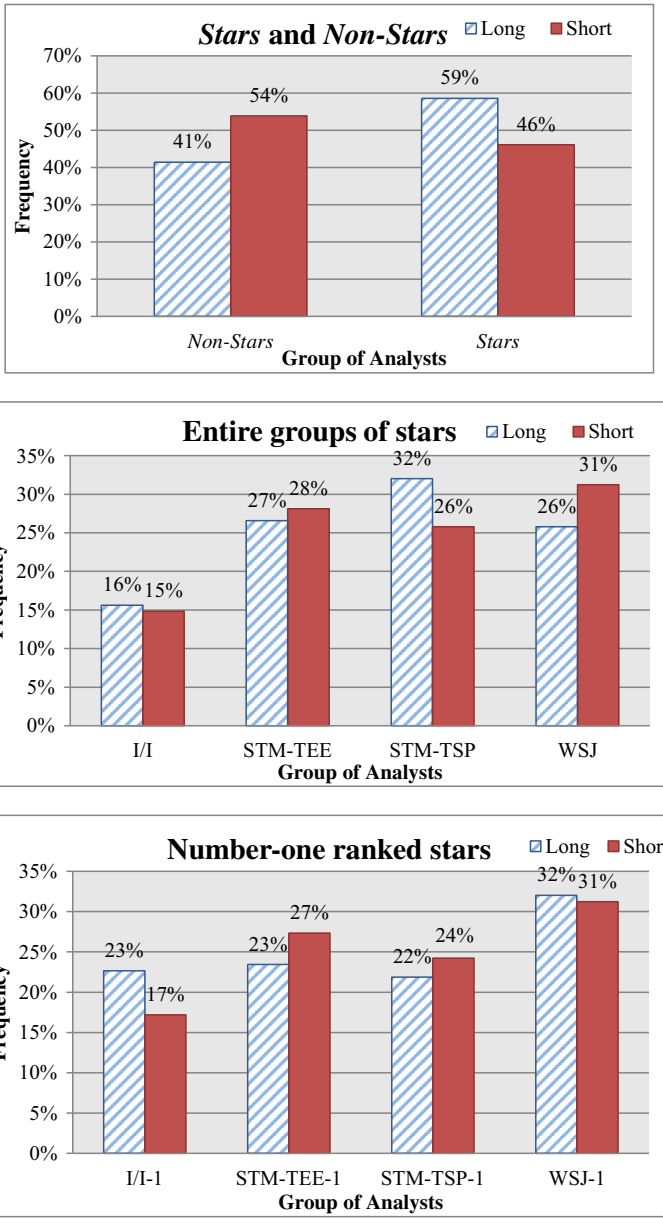

Fig. 1 Frequency of months when a particular group of analysts outperformed the other groups. Rankings by *Institutional Investor* (I/I), Thomson Reuters' *StarMine* "Top Earnings Estimators" (STM-TEE) and "Top Stock Pickers" (STM-TSP), and *The Wall Street Journal* (WSJ). Indexation by -1 signifies a group of number-one ranked analysts. Daily returns were geometrically aggregated to monthly values. Time period is from November 2003 until May 2014. Long portfolios contain Strong Buy and Buy recommendations; Short portfolios contain Sell and Strong Sell recommendations. Hold portfolios are not compared

An additional point to consider concerns conflicts of interest, which analysts face if they want to be selected by the subjective rankings, and which could reduce the value for investors

of the recommendations from stars selected by I/I ranking. Thus, analysts may issue over-optimistic recommendations on a stock that a particular buy-side client is overweighting. While this might lead to the poor performance of the analyst's stock picks, it might help to buy some votes in the coming I/I poll. This reasoning goes in line with the previous findings by Mola and Guidolin (2009), who reported that I/I star analysts are more prone to issue too-optimistic ratings for the stocks that affiliated mutual funds hold.

Among the objective rankings, we find significant differences in their predictive power, which we discuss as the persistence of the performance from the *Year Before* to the *Year After* election. We document a significant decrease in performance from the *Year Before* to the *Year After* for the analysts who were ranked according to the investment value of the recommendations (STM-TSP and WSJ). This result might be explained by the low predictive power of such recommendation-based election methods which lead to the regression to the mean, whereby the previous year's best performers should exhibit results that are closer to the average in subsequent years. However, we observed significantly positive returns for the Long portfolios of the STM-TSP and STM-TSP-1 stars, which outperformed *Non-Stars* in the year after election, even though there is a decline in performance compared with the evaluation year. Overall, our results confirm the previous findings that reported low predictive power for the methods which focus solely on the past performance of the recommendations (Emery and Li 2009).

In contrast to the results for objective recommendation-based rankings, for STM-TEE ranking that is based on the accuracy and timing of earnings forecasts, the returns for Long-Short portfolios in the *Year After* differ insignificantly from those of the *Year Before*. More important is that this group outperforms *Non-Stars* and some other groups of stars (I/I and STM-TPS) in the year after election. We document that this outperformance is explained by the high excess returns for the Short portfolio of STM-TEE, while their Long portfolio was insignificantly different from the *Non-Stars* as well as those of the other groups of stars. This conclusion emphasizes the importance of accurate earnings forecasts being used in the valuation models in order to predict future stock prices, especially the decrease in prices, which lead to the outperformance among Short portfolios.

Overall, we conclude that the stock-picking skill is difficult to capture by focusing only on the performance of the recommendations over a one-year horizon. However, methodologies which consider other fundamental skills, such as earnings forecasts that are necessary for successful stock picks, gave much higher predictive power for the performance of future recommendations, even though the previous year returns by our evaluation methodology for those earnings-based stars were not as high as for the recommendation-based rankings.

3.3 Robustness test

In this paper we employ a method used by other researchers (Barber et al. 2007; Fang and Yasuda 2014) holding the stocks in the portfolio for one year or until the next recommendation change from the same analyst, whichever happens earlier. Our baseline analysis considers all recommendation changes issued by all analysts listed in different rankings. In this section we discuss the results from imposing additional filters on the samples of recommendations and analysts. Also, we discuss alternative holding periods and construction of the Short portfolios.

First, we exclude the recommendation changes which were issued on the same days as earnings announcement days in order to avoid contamination in measuring analyst stock-

picking skill. While the previous study by Li et al. (2015) documents that the contemporaneous-with-news recommendation changes have higher market impact than the earnings announcements themselves since such recommendations help investors interpret the news, some of such recommendations might simply be piggybacking on the news without providing more value to investors. Excluding the recommendations that were issued on the same days as the earnings announcement dates obtained from the Compustat database (accounted for 3.4 % of recommendations), we find that our main conclusions remain the same, though the alphas were insignificantly decreased. Excluding the recommendations that were issued within a three-day window around earnings announcements (17.3 % of recommendations), we found slight changes in conclusions from our baseline analysis. Specifically, the Long portfolios of I/I in the *Year After* performed significantly better than the *Non-Stars*; also, the differences in returns for Long-Short portfolios among STM-TEE and STM-TSP became insignificant. However, the results of the comparison for the Long-Short portfolios of each group of stars with the *Non-Stars* remain unaffected: STM-TEE outperformed, while I/I underperformed the group of *Non-Stars*.

Second, we repeat our calculations using two filters on our sample of analysts: (i) non-overlapping sample of stars; and (ii) in each industry, limit the number of stars to one, two, and three only (applies to I/I and WSJ, while *StarMine* lists remain unchanged). While the non-overlapping samples exclude those analysts who are performing well according to some or even all evaluation methods (subjective or objective), such a limitation might help to distil the differences between rankings' performance. By considering only three stars in each industry in I/I and WSJ lists, we try to reduce the differences in the number of analysts in each group of stars and eliminate the influence of bottom-ranked analysts (runners-up) on our conclusions. In both cases, we focused on the differences among groups of stars and found that, in the case of focusing on only three stars per industry, our main conclusions remain the same. After examining the results for non-overlapping samples, we found that the highest change was for the groups of I/I and STM-TEE stars whose Long-Short portfolio returns were lower as compared to our baseline analysis. The decrease of the returns for some portfolios for non-overlapping groups of stars led to an insignificant difference in performance between groups of STM-TEE and I/I with STM-TSP stars. Since some star analysts appear in several rankings in a given year and the results are lower with non-overlapping samples of STM-TEE and I/I, we interpret this fact as an indication that the best analysts are selected as stars in several rankings at the same time.

Third, we use some alternative portfolio constructs: (i) we take the immediate year before rankings were announced instead of the previous calendar year; and (ii) we include Hold recommendations into Short portfolios, thus having only two portfolios to consider, namely Long and Short.

We take an immediate year before the rankings' announcement to use data that are more current and relevant. The fact that I/I and *StarMine* rankings are announced around October each year, and WSJ publishes its list of stars in May, potentially creates a bias, as information used for analysis is relatively stale for I/I and *StarMine* compared to WSJ. After we repeat our calculations for a new time-frame of the *Year Before*, we find that the excess returns for I/I changed insignificantly, while the returns for WSJ and STM-TSP decreased (especially for STM-TSP, since STM-TSP is published in October, the overlap of the immediate year before ranking is announced with the previous calendar year is lower than for WSJ, which is published in May). This decrease in the performance for recommendation-based rankings led to insignificant differences between the returns for Long-Short portfolios of STM-TSP and

WSJ with those of the STM-TEE in the *Year Before*. Also, the difference among Long-Short portfolios for STM-TSP and I/I stars became insignificant. These results are in line with our main findings as they show that high excess returns were documented only during the same year as the evaluation year, while they regressed to the mean in subsequent time periods.

We also tested alternative portfolio constructs by including Hold recommendations into the Short portfolio as in Barber et al. (2007) and in Fang and Yasuda (2014). Taking into account that analysts are reluctant to issue sell recommendations (Barber et al. 2007), investors might treat Hold ratings as a signal to sell, especially when this Hold is a downgrade from Buy or Strong Buy. While this might lead to a negative price reaction on some Hold recommendations, the upgrades from Sell/Strong Sell to Hold are expected to generate near market returns (Barber et al. 2010). We repeated our calculations and found that the alpha levels for Short portfolios that contain Hold/Sell/Strong Sell recommendations on a one-year horizon were close to zero for almost all portfolios.[6] That result is expected considering that such Short portfolios had about 90 % weight in Hold recommendations, and according to our results for "Long, Hold, Short" portfolio calculations, Hold portfolios have alphas insignificant from zero for all groups of analysts (except for the Hold portfolio of I/I stars, see Panel A in Table 7).

Additionally, we reconstructed all portfolios using shorter holding periods, such as 30 days as in (Fang and Yasuda 2014) and realized that the Short portfolios (Sell/Strong Sell recommendations) for some groups of stars had a very low number of stocks. Thus, Short portfolios of STM-TSP had only 9.1 stocks on average, STM-TEE had 7.5 stocks, and portfolios for the number-one ranked *StarMine* analysts had on average fewer than 5 stocks. We also tested other time periods within a horizon of one year (namely, 45, 60, 90, 180 days) and found that up to 180 days, the Short portfolios from the number-one ranked stars had a relatively low number of stocks (less than 10). Hence, those portfolios were highly undiversified and their excess returns were unreliable. Consequently, we decided to stick to a one-year horizon for considering the recommendations to be valid if they are unchanged, and we found that for short portfolios the number-one stars had on average 20 stocks and for the entire groups of stars the average number was more than 50 stocks during the investigated time period.

Finally, we tested combinations of different restrictions, such as a non-overlapping sample of only three stars in each industry, and excluding the recommendations on earnings announcement days. We find that all possible filters affect the alpha levels, but do not influence our main conclusions on the differences between groups of *Stars* and *Non-Stars*, while the differences among Long-Short portfolios for the groups of stars become statistically insignificant. All the results of the robustness tests remain unpublished and are available upon request from the authors.

4 Conclusion

The goal of this study was to determine whether star rankings can be employed as an indicator of the future profitability of analysts' recommendations. By using a unique database for the period from 2003 to 2014, we find that sell-side analysts indeed issue profitable

[6] Only the returns of the Short and Long-Short portfolios were affected in this robustness test, while the results for the Long portfolios remained unchanged.

recommendations. This conclusion is supported by the previous research of Mikhail et al. (2004), who find that sell-side analysts are persistent in issuing profitable recommendations. Our results challenge the finding by Emery and Li (2009) that star rankings are largely "popularity contests". In our study, we found that only Buy and Strong Buy recommendations from the entire group of Star analysts outperform those of the *Non-Stars* in the year after election, while Sell and Strong Sell recommendations performed as those of the *Non-Stars*. While *Stars* had significantly higher returns for their portfolios during the evaluation year, the group of *Stars* was not able to repeat the excess returns at the same level after rankings were announced.

After investigating the differences among subjective (I/I) and objective (*StarMine* and WSJ) rankings, we found that the returns of the recommendations from analysts ranked by the subjective ranking from I/I underperformed most of the other groups of stars as well as the group of *Non-Stars*. However, this finding does not contradict the intentions of the I/I ranking, whose methodology does not focus on the investment value of recommendations. Among objective rankings, we found that the most persistent results were observed for the group of STM-TEE analysts, who were selected based on the accuracy and timing of the earnings forecasts. Their recommendations outperformed the groups of stars from STM-TSP and I/I as well as *Non-Stars* in the year after election. We document that the recommendation-based STM-TSP and WSJ rankings show a significant decrease in performance from the evaluation year to the year after rankings were published, thus confirming the previous findings that reported a low predictive power for the methods which focus solely on the past performance of the recommendations (Emery and Li 2009).

In summary, the choice of which analysts to work with is of great importance for the long-term growth of an investor's portfolio. Our results show that the stock-picking skill is difficult to capture by focusing only on the performance of the recommendations over a one-year horizon. In our study, we provided empirical evidence regarding which star rankings of sell-side analysts a potential investor should have relied on, namely, the *StarMine*'s "Top Earnings Estimators". We find it comforting that estimation of future earnings is important for predicting portfolio returns, since valuation models used for valuing stocks are built on a company's future earnings. In conclusion, our results show that stock-picking ability reflects a set of skills that can be captured using more fundamental evaluation methods such as those that consider earnings forecasts.

Acknowledgments This research was conducted within the framework of the European Doctorate in Industrial Management (EDIM), which Yury O. Kucheev participates in, and was funded by The Education, Audiovisual and Culture Executive Agency of European Commission under Erasmus Mundus Action 1 program. We are grateful to the editor Haluk Ünal and two anonymous referees for insightful comments that have greatly improved this paper. We would like to thank Joaquin Ordieres from the Technical University of Madrid for help in data processing. Kathryn M. Kaminski made valuable contributions at the beginning of this project, when she was at the Stockholm School of Economics, which is gratefully acknowledged. We thank Per-Olof Edlund, Björn Hagströmer, Gustav Martinsson, Per Thulin, Gregg Vanourek, and seminar participants at KTH, Lund University, University of Gothenburg, and the Swedish House of Finance for valuable comments. An earlier draft of this paper has been circulated as "Star sell-side analysts listed by Institutional Investor, The Wall Street Journal, and StarMine. Whose recommendations are most profitable?"

All errors are the responsibility of the authors.

References

Bagnoli M, Watts SG, Zhang Y (2008) Reg-FD and the competitiveness of all-star analysts. J Account Public Policy 27:295–316. doi:10.1016/j.jaccpubpol.2008.06.002

Barber BM, Lehavy R, McNichols M, Trueman B (2001) Can investors profit from the prophets? Security analyst recommendations and stock returns. J Finance 56:531–563

Barber BM, Lehavy R, McNichols M, Trueman B (2006) Buys, holds, and sells: the distribution of investment banks' stock ratings and the implications for the profitability of analysts' recommendations. J Account Econ 41:87–117. doi:10.1016/j.jacceco.2005.10.001

Barber BM, Lehavy R, Trueman B (2007) Comparing the stock recommendation performance of investment banks and independent research firms. J Financ Econ 85:490–517. doi:10.1016/j.jfineco.2005.09.004

Barber BM, Lehavy R, Trueman B (2010) Ratings changes, ratings levels, and the predictive value of analysts' recommendations. Financ Manag 39:533–553

Beyer A, Guttman I (2011) The effect of trading volume on analysts' forecast bias. Account Rev 86:451–481. doi:10.2308/accr.00000030

Boni L, Womack KL (2006) Analysts, industries, and price momentum. J Financ Quant Anal 41:85–109

Carhart MM (1997) On persistence in mutual fund performance. J Finance 52:57–82

Emery DR, Li X (2009) Are the wall street analyst rankings popularity contests? J Financ Quant Anal 44:–411. doi:10.1017/S0022109009090140

Ertimur Y, Mayew WJ, Stubben SR (2011) Analyst reputation and the issuance of disaggregated earnings forecasts to I/B/E/S. Rev Account Stud 16:29–58. doi:10.1007/s11142–009–9116-5

Fang L, Yasuda A (2009) The effectiveness of reputation as a disciplinary mechanism in sell-side research. Rev Financ Stud 22:3735–3777. doi:10.1093/rfs/hhn116

Fang L, Yasuda A (2014) Are stars' opinions worth more? The relation between analyst reputation and recommendation values. J Financ Serv Res 46:235–269. doi:10.1007/s10693-013-0178-y

Kadan O, Madureira L, Wang R, Zach T (2009) Conflicts of interest and stock recommendations: the effects of the global settlement and related regulations. Rev Financ Stud 22:4189–4217. doi:10.1093/rfs/hhn109

Kerl A, Ohlert M (2015) Star-analysts' forecast accuracy and the role of corporate governance. J Financ Res 38: 93–120. doi:10.1111/jfir.12053

Kim MS, Zapatero F (2011) Competitive compensation and dispersion in analysts' recommendations. Available at SSRN 1569888

Leone A, Wu J (2007) What does it take to become a superstar? Evidence from institutional investor rankings of financial analysts. Evid Institutional Invest Rank Financ Anal May 23 2007 Simon Sch Bus Work Pap No FR 2–12.

Li EX, Ramesh K, Shen M, JS W (2015) Do analyst stock recommendations piggyback on recent corporate news? An analysis of regular-hour and after-hours revisions. J Account Res 53:821–861. doi:10.1111/1475-679X.12083

Loh RK (2010) Investor inattention and the underreaction to stock recommendations. Financ Manag 39:1223–1252

Loh RK, Mian GM (2006) Do accurate earnings forecasts facilitate superior investment recommendations? J Financ Econ 80:455–483

Loh RK, Stulz RM (2011) When are analyst recommendation changes influential? Rev Financ Stud 24:593–627. doi:10.1093/rfs/hhq094

Michaely R, Womack KL (1999) Conflict of interest and the credibility of underwriter analyst recommendations. Rev Financ Stud 12:653–686. doi:10.1093/rfs/12.4.653

Mikhail MB, Walther BR, Willis RH (2004) Do security analysts exhibit persistent differences in stock picking ability? J Financ Econ 74:67–91. doi:10.1016/j.jfineco.2003.11.001

Mola S, Guidolin M (2009) Affiliated mutual funds and analyst optimism. J Financ Econ 93:108–137. doi:10.1016/j.jfineco.2008.06.006

Ohlson JA (1995) Earnings, book values, and dividends in equity valuation. Contemp Account Res 11:661–687

StarMine (2015a) Thomson Reuters | About us. http://www.starmine.com/index.phtml?page_set=about_sm. Accessed 4 Feb 2015

StarMine (2015b) Thomson Reuters | The StarMine Analyst Awards. METHODOLOGY FOR THE UNITED STATES. http://www.starmineawards.com/methodology.php#filter. Accessed 10 Dec 2015

Stickel SE (1995) Anatomy of the performance of buy and sell recommendations. Financ Anal J 51:25–39

Womack KL (1996) Do brokerage analysts' recommendations have investment value? J Finance 51:137–167

Yaros JR, Imielinski T (2013) A Monte Carlo Measure to Improve Fairness in Equity Analyst Evaluation. Working Paper